PROBLEMS AND POTENTIAL OF AUTOCALIBRATING A HYDROLOGIC MODEL

M. W. Van Liew, J. G. Arnold, D. D. Bosch

ABSTRACT. *An investigation was conducted to evaluate strengths and limitations of manual calibration and the existing autocalibration tool in the watershed-scale model referred to as the Soil and Water Assessment Tool (SWAT). Performance of the model was tested on the Little River Experimental Watershed in Georgia and the Little Washita River Experimental Watershed in Oklahoma, both USDA-ARS watersheds. A long record of multi-gauge streamflow data on each of the watersheds was used for model calibration and validation. Model performance of the streamflow response in SWAT was assessed using a six-parameter manual calibration based on daily mass balance and visual inspection of hydrographs and duration of daily flow curves, a six-parameter autocalibration method based on the daily sum of squares of the residuals after ranking objective function (referred to as SSQRauto6), a six-parameter method based on the daily sum of squares of residuals (SSQauto6), and an eleven-parameter method based on the daily sum of square of residuals (SSQauto11). Results show that for both watersheds, manual calibration generally outperformed the autocalibration methods based on percent bias (PBIAS) and simulation of the range in magnitude of daily flows. For the calibration period on Little River subwatershed F, PBIAS was 0.0%, −24.0%, −21.5%, and +29.0% for the manual, SSQRauto6, SSQauto6, and SSQauto11 methods, respectively. Based on the coefficient of efficiency (NSE), the SSQauto6 and SSQauto11 methods gave substantially better results than manual calibration on the Little River watershed. On the Little Washita watershed, however, the manual approach generally outperformed the automated methods, based on the NSE error statistic. Results of this study suggest that the autocalibration option in SWAT provides a powerful, labor-saving tool that can be used to substantially reduce the frustration and uncertainty that often characterize manual calibrations. If used in combination with a manual approach, the autocalibration tool shows promising results in providing initial estimates for model parameters. To maintain mass balance and adequately represent the range in magnitude of output variables, manual adjustments may be necessary following autocalibration. Caution must also be exercised in utilizing the autocalibration tool so that the selection of initial lower and upper ranges in the parameters results in calibrated values that are representative of watershed conditions.*

Keywords. *Calibration, Hydrology, Modeling, Optimization, Simulation, SWAT, Validation.*

During the past few decades, the need to determine the impacts of variations in climate, climate change, and changes in land use or land management practices on water quality or water resources has resulted in the development of a number of hydrologic simulation models. Recent advances in computing capability and Geographical Information Systems (GIS) have led to increasingly sophisticated watershed-scale models that incorporate climatic, soils, topographic, and land use characteristics and are capable of addressing a host of issues related to water quality concerns, flood control, low flow management, and water availability. Notable examples of continuous watershed simulation models include the Hydrologic Simulation Program-Fortran (HSPF; Johanson et al., 1984), Soil and Water Assessment Tool (SWAT; Arnold et al., 1993), Precipitation Runoff Modeling System (PRMS; Leavesley et al., 1983), and Integrated Runoff Model-F Bultot (IRMB; Bultot and Dupriez, 1976). These conceptually based, watershed-scale models are computationally efficient, operate on a daily or sub-daily time step, and often lump hydrologic processes that occur over short time steps into simplified assumptions.

Because hydrologic simulation models are inexact representations that mimic the movement of water in the physical environment, they are incomplete in their description of both the elements and the processes present in that environment. They must therefore be calibrated to minimize the error between the output simulated by the model and the data collected in the field (Van Griensven, 2002). In order for a watershed-scale model to provide acceptable simulations of streamflow and water quality constituents, such as sediment, nutrients, and pesticides, observed data are necessary for model calibration. The calibration of a hydrologic model, especially a conceptual one, is complicated by the fact that values for a large number of parameters or coefficients must be estimated (Jacomino and Fields, 1997; Srinivasan et al., 1998; Motovilov et al., 1999; Carrubba, 2000). Before the onset of high-speed computers, most hydrologic models were calibrated exclusively in a "manual" (or expert) fashion,

Article was submitted for review in September 2004; approved for publication by the Soil & Water Division of ASAE in March 2005.

The authors are **Michael W. Van Liew, ASAE Member Engineer,** Research Hydrologist, USDA-ARS Grazinglands Research Lab, El Reno, Oklahoma; **Jeffrey G. Arnold,** Agricultural Engineer and Research Hydraulic Engineer, USDA-ARS Grassland Soil and Water Research Lab, Temple, Texas; and **David D. Bosch, ASAE Member Engineer,** Research Hydraulic Engineer and Civil Engineer, USDA-ARS Southeast Watershed Research Lab, Tifton, Georgia. **Corresponding author:** Michael W. Van Liew, USDA-ARS Grazinglands Research Lab, 7207 W. Cheyenne St., El Reno, OK 73036; phone: 405-262-5291; fax: 405-262-0133; e-mail: mvanliew@grl.ars.usda.gov.

whereby a hydrologist utilized knowledge of the watershed and experience with the model to adjust the parameters through a trial and error procedure, while visually comparing the observed and simulated outputs (Gupta et al., 1999). However, the increasing complexity of hydrologic models has made manual calibration a significantly more difficult task that is highly labor intensive and often characterized by a great deal of frustration. Because model structures are highly nonlinear, it is difficult for the hydrologic practitioner to know just how sensitive the different portions of the simulation outputs are to parameter adjustments (Gupta et al., 1999).

The various shortcomings associated with manual calibration methods have led to the development of automatic calibration techniques that utilize high-speed computers and efficient algorithms for matching model response to observed data. During the past three decades, a considerable amount of research has been conducted on the development and refinement of automated calibration approaches (Boyle et al., 2000). Sophisticated measures have been developed to evaluate the closeness of fit between simulated and observed outputs and to statistically analyze parameter uncertainty (Thiemann et al., 2001). Although an automated approach to model calibration may provide substantial savings in labor on the part of the modeler, the possibility exits whereby values of the calibrated parameters do not realistically reflect watershed characteristics. Care must therefore be taken in assigning appropriate lower and upper ranges in parameter values prior to initiating the autocalibration process. In addition, the calibrated parameter set obtained by an optimization scheme may not necessarily provide acceptable model simulations for certain ranges in the output variables. Boyle et al. (2000) point out that automatic model calibrations may not provide parameter estimates and hydrologic simulations that are considered acceptable by hydrologists responsible for operational streamflow forecasting.

In order to better understand the strengths and weaknesses of manual and automatic approaches to model calibration, we conducted an investigation to assess the potential of a recently developed autocalibration tool in a watershed-scale, conceptual model. The objective of the study was to determine the performance of automated calibration and manual calibration approaches in simulating streamflow using the Soil and Water Assessment Tool. Model tests for the two approaches were conducted on the USDA-ARS Little River Experimental Watershed in Georgia and the USDA-ARS Little Washita River Experimental Watershed in Oklahoma.

MODEL DESCRIPTION

The SWAT model was originally developed by the USDA ARS to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large ungauged basins (Arnold et al., 1998). SWAT incorporates features of several ARS models and is a direct outgrowth of the Simulator for Water Resources in Rural Basins (SWRRB; Williams et al., 1985). Specific models that contributed to the development of SWAT include Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS; Knisel, 1980), Groundwater Loading Effects on Agricultural Management Systems (GLEAMS; Leonard et al., 1987), and Erosion-Productivity Impact Calculator (EPIC; Williams et al., 1984). The USDA-SCS runoff curve number is used to estimate surface runoff from daily precipitation. The curve number is adjusted according to moisture conditions in the watershed (Arnold et al., 1993). SWAT can also be run on a sub-daily time step basis using the Green-Ampt infiltration method (Green and Ampt, 1911). Other hydrologic processes simulated by the model include evapotranspiration, infiltration, percolation losses, channel transmission losses, channel routing, and surface, lateral, shallow aquifer, and deep aquifer flow (Arnold and Allen, 1996). The runoff curve number option of SWAT (Neitsch et al., 2002) is adopted in this study. For both watersheds, the Hargreaves (1975) method was used to estimate potential evapotranspiration.

For modeling purposes in SWAT, a watershed is partitioned in a number of subbasins. Each subbasin delineated within the model is simulated as a homogenous area in terms of climatic conditions, but with additional subdivisions within each subbasin to represent different soils and land use types. Each of these subdivisions is referred to as a hydrologic response unit (HRU) and is assumed to be spatially uniform in terms of soils, land use, topographic and climatic data.

The 2003 version of SWAT includes a multi-objective, automated calibration procedure that was developed by Van Griensven and Bauwens (2003). The calibration procedure is based on the shuffled complex evolution algorithm (SCE-UA; Duan et al., 1992) that allows for the calibration of model parameters based on a single function. In a first step, the SCE-UA selects an initial population of parameters by random sampling throughout the feasible parameter space for p parameters to be optimized, based on given parameter ranges. The population is partitioned into several communities, each consisting of $2p + 1$ points. Each community is made to evolve based on a statistical “reproduction process” that uses the simplex method, an algorithm that evaluates the objective function in a systematic way with regard to the progress of the search in previous iterations (Nelder and Mead, 1965). At periodic stages in the evolution, the entire population is shuffled and points are reassigned to communities to ensure information sharing. As the search progresses, the entire population tends to converge toward the neighborhood of global optimization, provided the initial population size is sufficiently large (Duan et al., 1992). The SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing, and land surface modeling, and has generally been found to be robust, effective, and efficient (Duan, 2003).

In the optimization scheme developed for SWAT 2003, parameters in the model that affect hydrology or water quality can be changed in either a lumped (over the entire watershed) or distributed (for selected subbasins or HRUs) way. In addition, the parameters can be modified by replacement, by addition of an absolute change, or by multiplication of a relative change. Besides weight assignments for output variables that can be made in multi-objective calibrations (e.g., 50% streamflow, 30% sediment, 20% nutrients), the user can specify a particular objective function that is minimized. The objective function is an indicator of the deviation between a measured and a simulated series (Van Griensven and Bauwens, 2003). The following three objective functions are included in SWAT 2003:

$$SSQ = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(Q_{i,obs} - Q_{i,sim}\right)^2 \quad (1)$$

where

SSQ = sum of squares of the residuals

i = time series sequence of the measured and simulated pair

n = number of pairs of measured and simulated variables

$Q_{i,obs}$ = observed variable

$Q_{i,sim}$ = simulated variable.

Equation 1 represents the classical mean square error method that aims at matching a simulated time series to a measured series.

$$SSQR = \left(\frac{1}{n}\right)\sum_{j=1}^{n}\left(Q_{j,obs} - Q_{j,sim}\right)^2 \quad (2)$$

where

SSQR = sum of squares of the difference of the measured and simulated values after ranking

j = ranking sequence

$Q_{j,obs}$ = observed variable after ranking

$Q_{j,sim}$ = simulated variable after ranking

Equation 2 represents the fitting of the frequency distributions of the observed and simulated series. As opposed to the SSQ method, the time of occurrence of a given value of the variable is not accounted for in the SSQR method (Van Griensven and Bauwens, 2003).

$$TMC = (100)\, abs\left[\left(\frac{\sum_{i=1}^{n} Q_{i,sim}}{\sum_{i=1}^{n} Q_{i,obs}}\right) - 1\right] \quad (3)$$

where TMC is the total mass balance controller.

Equation 3 minimizes the error on the model bias, and calculates the deviation from the measured mass or volume and thus allows the mass balances to be controlled over the simulation period (Van Griensven and Bauwens, 2003). The TMC objective function might be particularly useful for hydrologic studies where accurate knowledge of the total flow volume is critical.

Test Watersheds

Little River Experimental Watershed

The 334 km^2 Little River Experimental Watershed is located in south central Georgia near Tifton (fig. 1). Climate in the region is characterized as humid subtropical with long, warm summers and short, mild winters, with an average

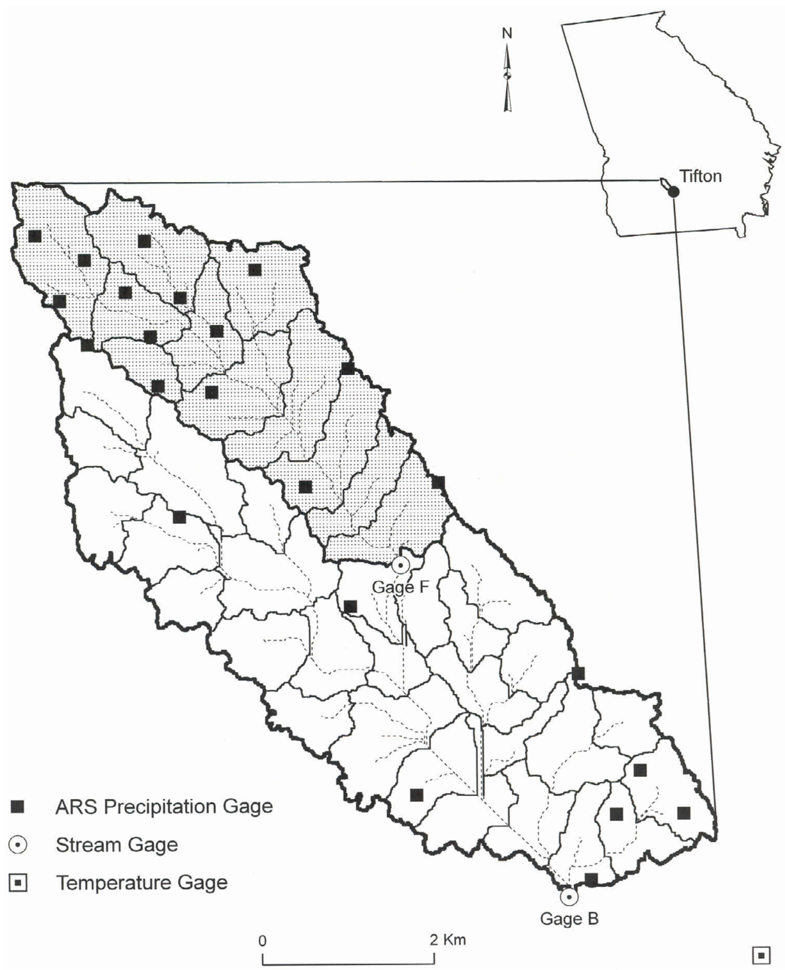

Figure 1. Location of the Little River Experimental Watershed in Georgia.

Table 1. Number of subbasins, number of hydrologic response units, drainage areas, and land use types for the two USDA ARS experimental watersheds.

Watershed	No. of Subbasins	No. of HRUs[a]	Drainage Area (km^2)	Land Use Type (%) Range/Pasture	Crop	Forest	Wetland	Misc.
Little River F	12	51	114	19	45	26	9	1
Little River B	40	161	330	10	42	45	2	1
Little Washita 526	22	138	160	59	28	6	0	7
Little Washita 550	73	486	600	66	19	9	0	6
Little Washita 522	66	413	538	66	18	9	0	7

[a] HRU = hydrologic response unit.

annual precipitation of about 1167 mm based on data collected by USDA ARS from 1971 to 2000. The Little River landscape is dominated by a dense dendritic network of stream channels bordered by riparian forest wetland and upland areas devoted mostly to agricultural uses. The region has low topographic relief and is characterized by broad, flat alluvial floodplains, river terraces, and gently sloping uplands (Sheridan, 1997). Upland soils on the watershed consist of fine-loamy to loamy siliceous, thermic Plinthic Paleudults. Bottomland soils are loamy, siliceous, thermic Arenic Plinthic Paleaquults with some Fluvaquents and Psammaquents (Lowrance et al., 1986). Since surface soils are underlain by shallow, relatively impermeable subsurface horizons, deep seepage and recharge to regional groundwater systems are impeded (Sheridan, 1997). Land use types include forest (45%), cropland (42%), rangeland and pasture (10%), wetland (2%), and miscellaneous (1%) (table 1). Almost year round production of vegetables and row crops such as peanuts and cotton has led to extensive and sustained use of fertilizer and pesticides on the watershed (Bosch et al., 2004). Increased animal production has elevated the risks associated with nonpoint-source pollution (Kellogg et al., 1994).

Upland areas between stream channels on the Little River are largely cultivated in pasture and cropland on small fields that are typically 20 ha in size. Planted pines occupy many of the upland and low-lying areas not in crops or pasture. Native vegetation is mostly swamp hardwoods in the bottomlands and pines with an understory of wire grass in the uplands. Approximately 20% of the upland fields are irrigated, primarily from agricultural ponds that dot the landscape.

Little Washita River Experimental Watershed

The Little Washita River Experimental Watershed is located about 80 km southwest of Oklahoma City and drains an area of 610 km^2 (fig. 2). The climate in the region is sub-humid to semi-arid, with an average annual precipitation of about 795 mm, based on data collected by USDA ARS from 1961 to 2000. Topography of the watershed is charac−

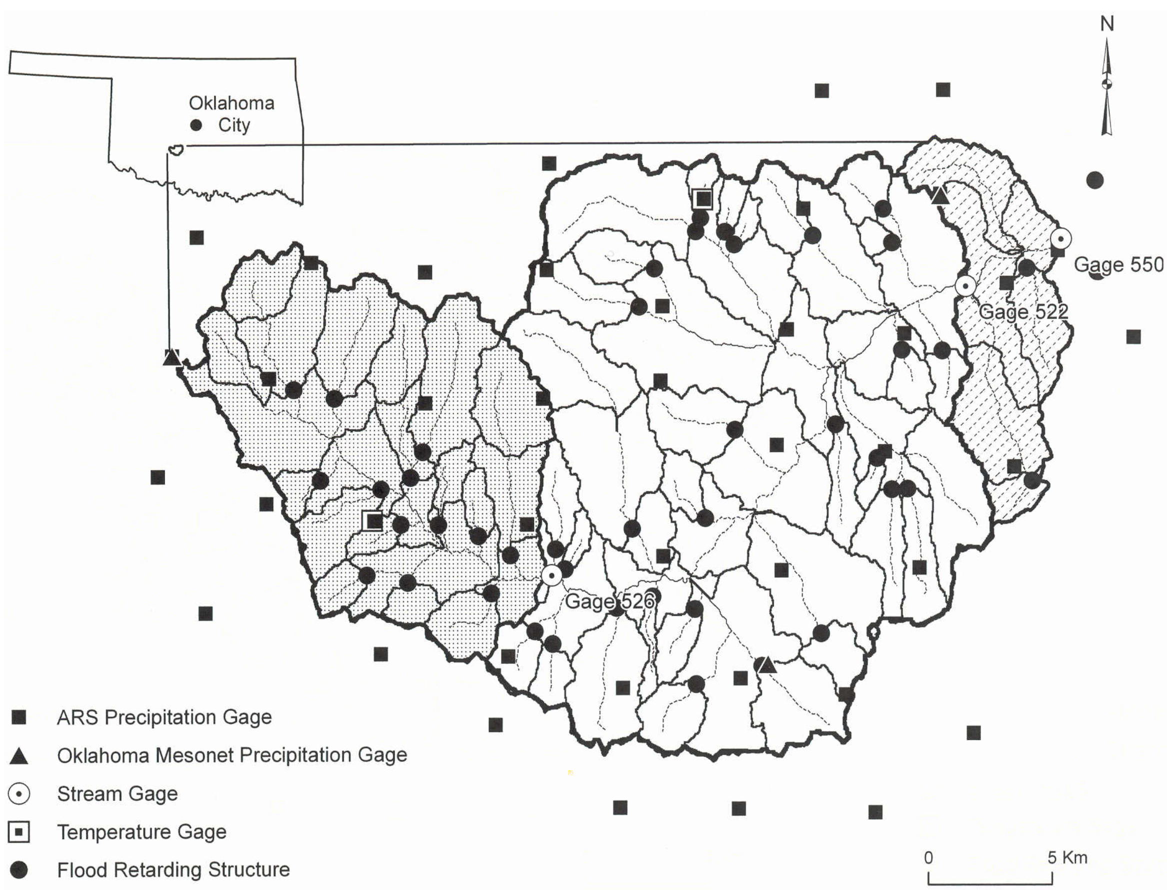

Figure 2. Location of the Little Washita River Experimental Watershed in Oklahoma.

terized by gently to moderately rolling hills. Soils include fine-silty, mixed thermic Udic and Pachic Argiustolls in the western portion; fine-loamy, siliceous thermic Ultic Haplustalfs in the southeastern portion; and coarse-silty, mixed, thermic Udic Haplustolls in the northeastern portion of the watershed. Bottomland soils along the floodplain consist of fine-silty mixed, thermic Cumulic Haplustolls (Water Quality and Watershed Research Laboratory, 1983). Land use types include rangeland and pasture (66%), cropland (18%), forest (9%), and miscellaneous (7%, including urban, abandoned oil fields, farmsteads, and ponds) (Allen and Naney, 1991).

Most crops that are grown on the Little Washita are cultivated using conventional tillage methods. Winter wheat is produced for both grain and the grazing of livestock during the winter months. Irrigation systems within the Little Washita are minimal, and occur mainly along the floodplain where alfalfa is raised. The watershed has numerous farm ponds located primarily in the lower portions of the watershed, and 45 USDA-NRCS flood-retarding structures (FRSs) constructed from 1969 to 1982. These FRSs control drainages on the Little Washita, range in size from 137 to 2860 ha, and consist of storage capacities ranging in size from 1.58×10^5 m^3 to 2.97×10^6 m^3. These structures delay and reduce peak surface flows and modify subsurface flows. They also lead to small increases in average annual evaporation due to a larger percentage of the watershed existing as a free water surface.

MODEL CALIBRATION

CALIBRATION PARAMETERS

Eleven calibration parameters in SWAT were selected that govern rainfall runoff processes on the two test watersheds. Model parameters were grouped into three categories, which were considered to predominantly govern surface, subsurface, and basin response, as shown in table 2. Following is a brief description of each parameter.

Calibration parameters governing the surface water response in SWAT include the runoff curve number, the soil evaporation compensation factor, and the available soil water capacity. The runoff curve number (CN2) is used to compute runoff depth as a function of total rainfall depth. It is a function of watershed properties that include soil type, land use and treatment, ground surface condition, and antecedent moisture condition. The soil evaporation compensation factor (ESCO) adjusts the depth distribution for evaporation from the soil to account for the effect of capillary action, crusting, and cracks. The available soil water capacity (SOL_AWC) is the volume of water that is available to plants at field capacity. It is estimated by determining the amount of water released between in-situ field capacity and permanent wilting point.

Six calibration parameters govern the subsurface water response in SWAT. One of these parameters is referred to as the groundwater “revap” coefficient (GW_REVAP), which controls the amount of water that will move from the shallow aquifer to the root zone as a result of soil moisture depletion and the amount of direct groundwater uptake from deep-rooted trees and shrubs. Another parameter that governs the subsurface response is the threshold depth of water in the shallow aquifer for “revap” to occur (REVAPMN). Movement of water from the shallow aquifer to the root zone or to plants is allowed only if the depth of water in the shallow aquifer is equal to or greater than the minimum “revap.” A third parameter is the threshold depth of water in the shallow aquifer required for return flow to occur to the stream (GWQMN). Two other parameters that govern watershed response include the baseflow alpha factor and groundwater delay. The baseflow alpha factor (ALPHA_BF), or recession constant, characterizes the groundwater recession curve. This factor approaches one for flat recessions and approaches zero for steep recessions. The groundwater delay (GW_DELAY) is the time required for water leaving the bottom of the root zone to reach the shallow aquifer. A sixth factor is the deep aquifer percolation fraction, which governs the fraction of percolation from the root zone to the deep aquifer (RCHRG_DP).

The surface runoff lag time (SURLAG) in SWAT provides a storage factor in the model for surface runoff to allow runoff to take longer than one day to reach a subbasin outlet. Channel hydraulic conductivity (CH_K2) governs the movement of water from the streambed to the subsurface for ephemeral or transient streams.

Table 2. Listing of parameters, their description, and respective units that were calibrated in SWAT.

Parameter	Description	Units
Parameters governing surface water response		
CN2	SCS runoff curve number	none
ESCO[a]	Soil evaporation compensation factor	none
SOL_AWC[a]	Available soil water capacity	mm mm^{-1}
Parameters governing subsurface water response		
GW_REVAP	Groundwater “revap” coefficient	none
REVAPMN	Threshold depths of water in the shallow aquifer required for “revap” to occur	mm
GWQMN[a]	Threshold depths of water in the shallow aquifer required for return flow to occur	mm
GW_DELAY	Groundwater delay	days
ALPHA_BF	Baseflow alpha factor, or recession constant	days
RCHRG_DP	Deep aquifer percolation fraction	fraction
Parameters governing basin response		
SURLAG[a]	Surface runoff lag time	days
CH_K2[a]	Channel hydraulic conductivity	mm h^{-1}

[a] Parameters calibrated in the SSQauto11 method; default values assigned for all other methods.

Watershed Delineation and Data Input

Subbasins were delineated in the Little Washita to account for variations in rainfall based on the spatial distribution of precipitation gauges and for the impact of the FRSs on hydrologic response in the watershed. The resulting subbasin density in the watershed was approximately 12 subbasins per 100 km^2. Preliminary testing on the Little Washita showed that further delineation of the watershed into more subbasins did not improve simulation results. For consistency in model simulations between the two watersheds, the approximate subbasin density used in the Little Washita was also used in the Little River. Delineation of subbasins for the two watersheds is illustrated in figures 1 and 2, along with the location of precipitation and stream gauges used in the analysis.

For this study, the number of HRUs in SWAT was based on a land use and soil type covering an area of at least 5% and 20%, respectively, within any given subbasin. Although smaller percentage thresholds for land use and soil type could have been selected, the additional computational time that would have been required for model simulations was not warranted. Table 1 lists the respective number of subbasins and HRUs delineated for each of the simulated subwatersheds. Although there have been minor changes in land use types over time on both the Little River and Little Washita River watersheds, records are not available on either watershed to accurately denote year to year changes that have occurred. Moreover, preliminary testing conducted with SWAT on the Little Washita showed that changes in land use as indicated by available records resulted in only very minor changes in streamflow. For this study, it was therefore assumed that the respective land use types on each of the watersheds remained constant for the period of record simulated.

Hydrologic data collected from 1997 to 2002 were used to calibrate subwatersheds F and B on the Little River Experimental Watershed (fig. 1). This was the period of record selected for manual calibration in a previous study by Bosch et al. (2004). Streamflow response for the Little River calibration period consists of two very wet years (1997-1998 calendar years) followed by four very dry years that were punctuated by four consecutive periods of negligible streamflow (Sept.-Nov. 1999; June-Aug. 2000; Aug.-Nov. 2001; and June-Sept. 2002). An earlier record of hydrologic data on the watershed from 1972 to 1996 was used for model validation.

Available hydrologic data from 1993 to 1999 were used to calibrate SWAT on subwatersheds 526 and 550 of the Little Washita (fig. 2). This period of record corresponded to a similar period used by Van Liew and Garbrecht (2003) for manual calibration of the model. The average annual precipitation measured by the USDA ARS for this 7-year period was 832 mm. Hydrologic data collected on subwatershed 526 from 1980 to 1985 and on subwatershed 522 from 1963 to 1985 were in turn used for model validation. To determine the robustness of the calibration methods to simulate streamflow conditions under different climatic conditions, the hydrologic record on subwatershed 522 was divided into three 5-year periods: a much dryer (697 mm) than average period (1964 to 1968), a near-average (765 mm) period (1975 to 1979), and a wetter (888 mm) than average period (1981 to 1985). These three periods were chosen from the available historical record, based on 5-year averages of precipitation from 1963 to 1985. Using a 40-year precipitation average on the Little Washita equal to 795 mm, the much dryer, near-average, and wetter than average periods specified on subwatershed 522 represent departures from the norm of about −12%, +4%, and +12%, respectively.

Calibration Approaches

Three types of calibrations were performed in this study, including a manual calibration, an autocalibration, and an expanded autocalibration. The manual calibration was first conducted at the monthly time step followed by a refined calibration at the daily time step. Each type of autocalibration was performed at the daily time step. Due to the complexity of manually calibrating a multitude of model parameters in SWAT, the manual approach to model calibration was limited to six parameters. These six parameters were also used in a previous study on the Little Washita that involved a manual calibration (Van Liew and Garbrecht, 2003). Using SWAT 2000, these six parameters included one surface water parameter (CN2) and five subsurface water parameters (GW_REVAP, REVAPMN, ALPHA_BF, GW_DELAY, and RCHRG_DP). These same six parameters were in turn calibrated using the autocalibration tool in SWAT 2003. In the expanded autocalibration, all eleven model parameters described above and listed in table 2 were calibrated using SWAT 2003. Although streamflow modeled in SWAT is comprised of surface, lateral, and subsurface flow elements, only the total streamflow component was compared in the investigation.

Manual Calibration

SWAT was calibrated manually by following a multistep procedure recommended by Neitsch et al. (2002). For the Little Washita, the upper watershed (subwatershed 526) was calibrated first, and the parameters in that subwatershed were then held constant while the larger watershed (subwatershed 550) was calibrated on that portion of the subwatershed below the outlet of 526. Although it was recognized that computational differences between measured and simulated streamflow at the outlet of subwatershed 526 were passed onto subwatershed 550, this approach to calibration was the most reasonable option that could be exercised, based on the availability and quality of existing data sets in the watershed. A similar approach was taken in calibrating the Little River. The upper portion of the watershed (subwatershed F) was calibrated first, which was then followed by a calibration of subwatershed B. The TMC method (eq. 3) was used as the optimization scheme for the manual calibration on a daily basis, accompanied by visual inspection of daily hydrographs and duration of daily flow curves. The sum of squares of residuals SSQ objective function (eq. 1) could have been used in the manual calibration, but preliminary testing showed that the TMC method in combination with the inspection of duration of daily flow curves gave a better representation of the range of simulated flows.

Detailed calibration of SWAT on the Little Washita was previously reported by Van Liew and Garbrecht (2003). A preliminary calibration was conducted on a monthly basis to identify the order of magnitude of all parameters to reproduce proper runoff volumes and seasonal characteristics. Briefly, the runoff curve number (CN2) that governs the surface runoff response was first calibrated. Second, the groundwater "revap" coefficient (GW_REVAP), the threshold depth of water in the shallow aquifer required for return flow (REVAPMN), and the deep aquifer percolation fraction (RCHRG_DP), which governs the fraction of percolation from the root zone to the deep aquifer, were calibrated. Third, the baseflow recession factor (ALPHA_BF) and the groundwater delay (GW_DELAY) parameter were calibrated so that the monthly measured versus simulated hydrographs agreed well. This preliminary calibration was followed by a fine-tuning at the daily time scale so that the predicted versus measured peak flows and recession curves on a daily time step matched as closely as possible. This same approach was taken in the manual calibration of the Little River.

Autocalibration

Parameters in SWAT were calibrated at the daily time scale in a distributed fashion using the automated calibration procedure, where observed and simulated outputs were compared at the same outlet points on the watersheds as described for the manual calibration. With the completion of a given optimization, two sets of calibrated parameters were computed for the Little River that corresponded to subwatersheds F and B, and two sets were computed for the Little Washita that corresponded to subwatersheds 526 and 550. Default values suggested by Van Griensven (2002) were selected as initial upper and lower ranges for the respective model parameters. The same six parameters used in the manual calibration were calibrated using the autocalibration tool in the model. One six-parameter autocalibration consisted of using the SSQ objective function (eq. 1), herein referred to as SSQauto6. A second six-parameter autocalibration consisted of using the SSQR objective function (eq. 2), and is referred to as SSQRauto6. A third six-parameter autocalibration consisted of using the TMC objective function (eq. 3). However, preliminary testing with this method yielded such poor simulation results that it was eliminated as one of the autocalibration approaches. An expanded autocalibration for comparing measured versus simulated streamflow included all eleven parameters using the SSQ objective function (eq. 1), and is referred to as the SSQauto11 approach.

EVALUATION CRITIERIA

Four evaluation criteria were used to assess monthly and daily streamflow simulated by SWAT. The first two criteria were error statistics that quantitatively measured the agreement between simulated and observed values, and the second two criteria involved visual comparisons of plots of simulated and observed values. The first evaluation criterion used was the percent bias (PBIAS), which is a measure of the average tendency of the simulated flows to be larger or smaller than their observed values. The optimal PBIAS value is 0.0; a positive value indicates a model bias toward underestimation, whereas a negative value indicates a bias toward overestimation (Gutpa et al., 1999). PBIAS may be expressed as:

$$\text{PBIAS} = \frac{\sum_{i=1}^{n}\left(Q_{i,obs} - Q_{i,sim}\right)(100)}{\sum_{i=1}^{n}\left(Q_{i,obs}\right)} \quad (4)$$

where

PBIAS = deviation of streamflow discharge, expressed as a percent

$Q_{i,obs}$ = observed streamflow in cubic meters per second

$Q_{i,sim}$ = simulated streamflow in cubic meters per second.

For purposes of comparison, three ranges in PBIAS values were arbitrarily chosen for this study. Computed values of PBIAS less than ±20% were considered good, values between ±20% and ±40% were considered satisfactory, and those greater than ±40% were considered unsatisfactory.

The second evaluation criterion was the model coefficient of efficiency (Nash and Sutcliffe, 1970), which expresses the fraction of the measured streamflow variance that is reproduced by the model. In this study, the Nash-Sutcliffe coefficient of efficiency (NSE) was computed for both monthly and daily output. For monthly streamflow, for example, model efficiency may be expressed as:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{i,obs} - Q_{i,sim}\right)^2}{\sum_{i=1}^{n}\left(Q_{i,obs} - Q_{mean}\right)^2} \quad (5)$$

where

NSE = coefficient of efficiency

Q_{mean} = mean observed streamflow during the evaluation period in cubic meters per second.

Comparison of equations 1 and 5 shows that both expressions contain:

$$\sum_{i=1}^{n}\left(Q_{i\,obs} - Q_{i\,sim}\right)^2$$

Minimizing the difference between $Q_{i,obs}$ and $Q_{i,sim}$ in equation 1 through optimization results in maximizing NSE for a given time series that is calibrated. Using NSE to evaluate the SSQ and SSQR objective functions given by equations 1 and 2, respectively, therefore favors the SSQ method. For this study, simulation results were considered to be good for values of NSE greater than 0.75, while for values of efficiency between 0.75 and 0.36, the simulation results were considered to be satisfactory (Motovilov et al., 1999). NSE values less than 0.36 were considered to be unsatisfactory.

The third criterion compared simulated monthly and daily hydrographs to observed values. At the daily time scale, particular attention was given to the timing and magnitude of peak flows and the shape of recession curves. The fourth criterion compared measured versus simulated flow duration curves to determine how well the model predicted the range in magnitude of daily flows throughout the calibration or validation periods.

The four criteria mentioned above were used to describe differences between measured and simulated hydrologic responses for the two watersheds. As noted by Jacomino and Fields (1997), possible explanations for differences between measured and simulated results may be classified as either "data deficiencies" or "model deficiencies." Data deficiencies refer to the use of erroneous watershed data that are recorded at the watershed but may apply to only a portion of it. Model deficiencies refer to simplifications or limitations that exist in the model that prevent it from adequately accounting for physical processes that occur on the watershed. Van Liew and Garbrecht (2003) provide examples of data and model deficiencies related to SWAT's calibration on

Table 3a. Parameter values calibrated in SWAT 2000 for the manual approach and SWAT 2003 for the automated approach.

Watershed	Parameter	Initial Default Values Lower Bound	Initial Default Values Upper Bound	Calibration Method Manual		SSQRauto6		SSQauto6		SSQauto11	
Little River				F	B	F	B	F	B	F	B
	CN2[a]	−50	50	−27	−27	−27	−50	−50	−50	−48	−46
	GW_REVAP	0.02	0.2	0.14	0.2	0.06	0.02	0.17	0.18	0.12	0.03
	REVAPMN	0	500	3	0	392	0	450	0.18	46	449
	GW_DELAY	0	500	2	2	0	0	9.1	5	2.5	2.1
	ALPHA_BF	0	1	1	1	0.21	0.89	0.87	0.36	0.7	0.3
	RCHRG_DP	0	1	0.05	0.12	0.18	1	0.04	0.24	0.92	0.01
	SURLAG[b]	0.5	10	4	4	4	4	4	4	0.53	0.53
	CH_K2[b]	0	150	0	0	0	0	0	0	116	16.6
	GWQMN[b]	0	5000	0	0	0	0	0	0	26.9	22.8
	ESCO[b]	0	1	0.95	0.95	0.95	0.95	0.95	0.95	0.13	0.34
	SOL_AWC[a],[b]	−50	50	0	0	0	0	0	0	−17	22.9
Little Washita				526	550	526	550	526	550	526	550
	CN2[a]	−50	50	−13	−13	−22	−2.7	−47	−13	−36	13
	GW_REVAP	0.02	0.2	0.03	0.13	0.12	0.09	0.09	0.13	0.09	0.18
	REVAPMN	0	500	0	0	0.02	0.07	0.01	0.05	471	192
	GW_DELAY	0	500	280	380	48.4	33	3.5	42.7	105	211
	ALPHA_BF	0	1	1	1	0.12	0.26	0.07	0.47	0.79	0.92
	RCHRG_DP	0	1	0.05	0.25	0.71	0.89	0.68	0.52	0.27	0.58
	SURLAG[b]	0.5	10	4	4	4	4	4	4	0.72	0.72
	CH_K2[b]	0	150	0	0	0	0	0	0	121	101
	GWQMN[b]	0	5000	0	0	0	0	0	0	192	3560
	ESCO[b]	0	1	0.95	0.95	0.95	0.95	0.95	0.95	0.31	0.89
	SOL_AWC[a],[b]	−50	50	0	0	0	0	0	0	0.3	21

[a] Parameter values expressed as percent change from default values.
[b] Parameters calibrated in the SSQauto11 method; default values assigned for all other methods.

the Little Washita in a previous investigation. Recognition of these types of deficiencies was helpful in better understanding simulation differences that resulted among calibration approaches used in this study.

Results

Table 3a is a compilation of parameter values that were calibrated with the manual and the three autocalibration approaches for the two watersheds. Also shown in the table are the initial lower and upper bounds for parameter values that were user specified for this study. Values of the SCS runoff curve number (CN2) and available soil water capacity (SOL_AWC) in table 3a are expressed as a percent change from the default values. Since these two parameters were calibrated for each HRU, the calibrated data set consisted of a multitude of values for these parameters for each watershed. For brevity, a complete listing of these individual values of CN2 and SOL_AWC are not reported in the study. For illustrative purposes, however, default and calibrated values of SOL_AWC for each soil type on Little River subwatershed B are presented in table 3b for the SSQauto11 calibration approach. Default and calibrated values of CN2 for the various land cover types on the Tifton soil are also shown in the table.

Perhaps the most noticeable observation that is apparent in tables 3a and 3b from both the manual calibration and the autocalibrations is the percent departures in the calibrated values of the SCS curve number. These departures were more pronounced on the Little River watershed, where they ranged from −27% for the manual approach to −50% for the SSQRauto6 and SSQauto6 approaches. Implications regarding the calibrated values for CN2 are presented later in the Discussion section of this article. Other than the groundwater delay parameter, very few trends could be discerned regarding the magnitude of the parameters calibrated by the various approaches. For the Little River watershed, calibrated values of GW_DELAY in all cases were less than 10 days, indicating that a very short time was required for water to leave the root zone and reach the watershed's shallow aquifer. This would indeed be consistent with our understanding of the subsurface movement of water on that watershed. With one exception, calibrated values of GW_DELAY were greater than 30 days for Little Washita. Such values are more charac−

Table 3b. Calibrated values of SOL_AWC for each soil type and CN2 for various land cover types on the Tifton soil for Little River subwatershed B using the SSQauto11 calibration approach.

Soil Type	SOL_AWC Default Value	SOL_AWC Calibrated Value
Tifton	217	267
Osier	111	136
Pelham	186	229
Troup	143	176

Land Cover Type on Tifton Soil	CN2 Default Value	CN2 Calibrated Value
Pasture	69	37
Forested wetlands	66	36
Deciduous forest	66	36
Evergreen forest	55	30
Mixed forest	60	32
Agricultural crops	77	42

Table 4. Average annual measured and simulated streamflow, percent bias (PBIAS), and monthly and daily coefficients of efficiency for the manual and autocalibration approaches.

			Meas.	Manual[a]				SSQRauto6[a]				SSQauto6[a]				SSQauto11[b]			
Watershed	Area $(km)^2$	Time Series	Q (mm)	SQ (mm)	PBIAS (%)	M NSE	D NSE	SQ (mm)	PBIAS (%)	M NSE	D NSE	SQ (mm)	PBIAS (%)	M NSE	D NSE	SQ (mm)	PBIAS (%)	M NSE	D NSE
Calibration runs																			
LR F	113.5	1997-2002	310	310	0	0.44	0.18	385	−24	0.74	0.13	377	−21.5	0.82	0.57	220	29	0.82	0.7
LR B	330	1997-2002	271	270	−0.2	0.47	−0.36	219	19.2	0.58	−0.2	317	−17.2	0.81	0.63	226	16.4	0.92	0.78
Validation runs																			
LR F	113.5	1972-1996	374	408	−9	0.5	−0.13	506	−35.3	0.7	0.05	501	−34.1	0.78	0.55	304	18.7	0.81	0.65
LR B	330	1972-1996	348	337	3.2	0.47	−0.16	285	18.1	0.65	−0.25	413	−18.6	0.79	0.62	348	−0.9	0.9	0.68
Calibration runs																			
LW 526	160	1993-1999	153	153	0	0.87	0.6	149	2.6	0.07	0.17	162	−6.4	0.66	0.39	162	−5.9	0.6	0.34
LW 550	600	1993-1999	118	117	0.5	0.75	0.57	116	1.4	0.33	0.1	117	0.9	0.71	0.55	98	16.9	0.78	0.57
Validation runs																			
LW 526	160	1980-1985	103	116	−12.2	0.87	0.74	128	−24.6	0.58	0.14	143	−38.5	0.69	0.65	136	−32	0.25	0.36
LW 522	538	1964-1968	29	28	1.9	0.71	0.38	78	−171.6	−2.64	−0.98	68	−136.2	−2.61	−0.04	35	−20.2	0.44	0.3
LW 522	538	1975-1979	65	60	8	0.89	0.41	104	−61.2	0.22	−0.69	92	−42	0.67	0.29	82	−27.2	0.72	0.43
LW 522	538	1981-1985	97	139	−43.9	0.2	0.12	138	−42.3	0.04	−0.65	164	−68.9	0.03	0.27	131	−35	0.51	0.6

[a] 6-parameter calibration of CN2, GW_REVAP, REVAPMN, ALPHA_BF, GW_DELAY, and RCHRG_DP.
[b] 11-parameter calibration of CN2, GW_REVAP, REVAPMN, ALPHA_BF, GW_DELAY, RCHRG_DP, ESCO, SOL_AWC, GWQMN, CH_K2, and SURLAG.
SSQ = sum of squares of residuals.
SSQR = sum of squares of the difference of the measured and simulated values after ranking.
SQ = simulated streamflow (mm).
PBIAS = percent bias (%)
M NSE = monthly Nash-Sutcliffe coefficient of efficiency.
D NSE = daily Nash-Sutcliffe coefficient of efficiency.

teristic of a deeper groundwater aquifer that is present on the Little Washita in comparison to the Little River watershed.

Calibration and Validation Performance on the Little River Experimental Watershed

Average annual observed and simulated streamflow values for the manual and three autocalibration approaches for the respective calibration and validation periods on the two watersheds are presented in table 4. The table also lists the computed values of the percent bias (eq. 4) and coefficient of efficiency (eq. 5) on a monthly and daily basis for each data set. Test results were considered good (PBIAS < ±20%) for the manual calibration on Little River subwatershed F, satisfactory (±20% ≤ PBIAS ≤ ±40%) for the three autocalibration approaches on subwatershed F, and good for all four approaches on subwatershed B for both the calibration and validation data sets. Monthly NSE values for the calibration and validation periods on both Little River subwatersheds were considered good for the SSQauto11 and SSQauto6 approaches and satisfactory for the manual and SSQRauto6 methods. Based on NSE values at the daily time scale for the calibration and validation periods, SSQauto6 and SSQauto11 clearly outperformed the other two approaches that yielded unsatisfactory values.

Figure 3 displays monthly streamflow for the four calibration approaches on Little River subwatershed F for the calibration period from 1997 to 2002. The figure shows that each method used in calibration underestimated peak flow months that occurred during February 1997 and December 1997 to March 1998. Although a data deficiency could be the cause for this discrepancy between measured and simulated results, it is more likely the result of a deficiency in the model. A comparison of calibration methods suggests that while the SSQRauto6 and SSQauto6 approaches tended to overestimate low flows, the manual approach not only overestimated months with low flows, but underestimated peak months as well. Results of the SSQauto11 simulation show that even though this method underestimated streamflow amount by 29%, it more closely simulated the monthly peaks and baseflows than did other three methods.

Differences in hydrograph simulation were more pronounced in the comparison of daily streamflow on subwatershed F for the period of record from 1 January 1997 to 30 June 1997 (fig. 4). Results show that the manual and SSQRauto6 approaches did about equally well simulating the magnitude of storm peaks, while the SSQauto6 approach tended to underestimate peak flows but more closely simulate measured recession curves. The SSQauto11 calibration appeared to provide the best overall fit of streamflow response to storm events, which was mainly attributed to the inclusion of the SURLAG, SOL_AWC, and ESCO parameters in the calibration. Similar hydrograph shapes shown in figure 4 were also observed in comparing calibration results obtained for Little River subwatershed B. Each calibration approach on both subwatersheds estimated the time to peak one day too early, a problem encountered with the SWAT model by Bosch et al. (2004) in a previous investigation on the Little River.

Figure 5 illustrates duration of daily flow for Little River subwatershed B for the calibration period from 1997 to 2002, and is indicative of how well the manual and autocalibration approaches simulated the range in magnitude of daily flows. The figure shows that although each of the calibration schemes did a reasonable job of simulating the upper range in flows, the manual approach outperformed each of the autocalibration approaches in the lower range in flows. Plots of daily mean flow duration for the validation period on Little River subwatershed B and both the calibration and validation periods on subwatershed F yielded results that were similar to those shown in figure 5.

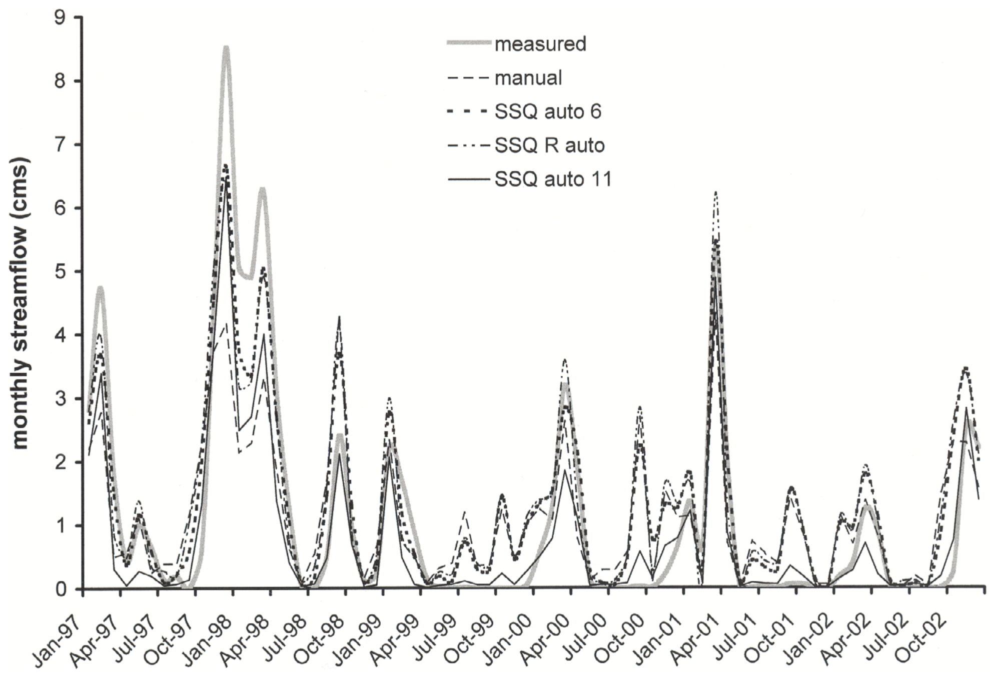

Figure 3. Comparison of monthly streamflow on Little River subwatershed F for the calibration period from 1997 to 2002.

CALIBRATION AND VALIDATION PERFORMANCE ON THE LITTLE WASHITA RIVER EXPERIMENTAL WATERSHED

The percent bias for all calibration approaches compared very well for the calibration period on Little Washita subwatersheds 526 and 550 (table 4). For the validation period on subwatershed 526, however, each calibration method overestimated streamflow amounts, ranging from 12.2% for manual calibration to 38.5% for SSQRauto6. For the dry, average, and wet validation periods on subwatershed 522, the manual approach gave much better results in estimating percent bias than did the autocalibration methods.

For both the calibration and validation periods on Little Washita River subwatershed 526, the manual calibration outperformed each of the autocalibration methods, based on monthly and daily values of the coefficient of efficiency. No satisfactory explanation could be given as to why the manual calibration was substantially better the autocalibrations obtained on this subwatershed. For the calibration period on subwatershed 550, test results show that the NSE values computed with the manual approach were comparable to those obtained for the SSQauto6 and SSQauto11 methods.

Simulation results also show that for the dry and average validation periods on subwatershed 522, the manual calibration outperformed each of the autocalibration methods, based on the coefficient of efficiency. For the wet climatic validation period, the SSQauto11 approach gave satisfactory

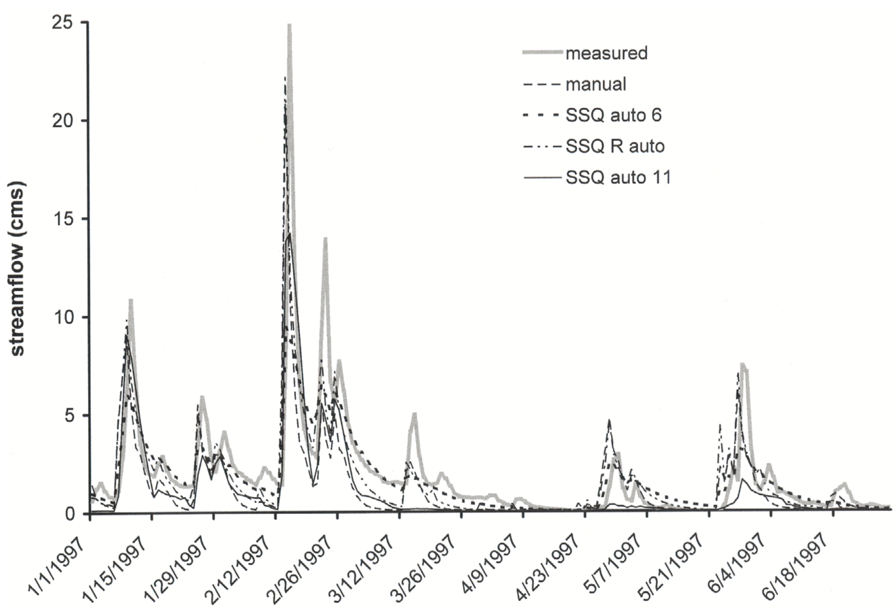

Figure 4. Comparison of daily streamflow on Little River subwatershed F for 1 January 1997 to 30 June 1997.

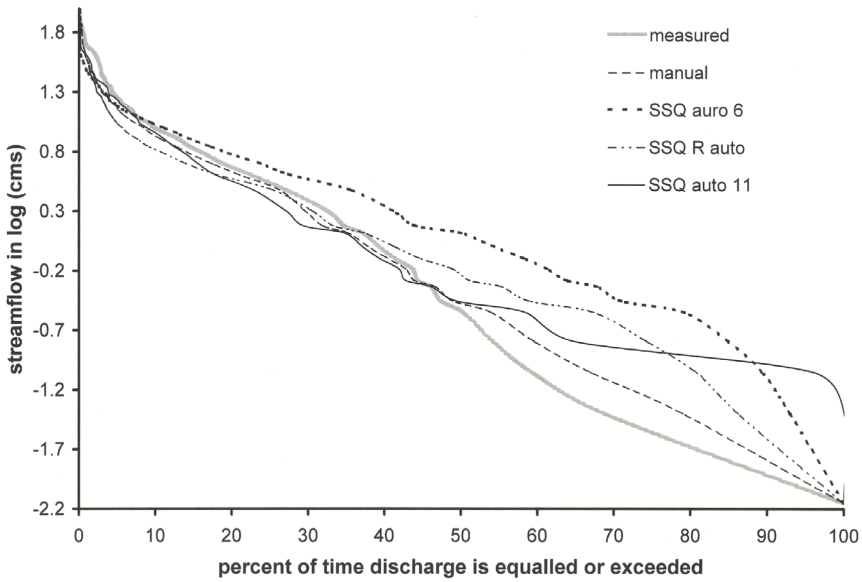

Figure 5. Duration of daily flow for Little River subwatershed B for the calibration period from 1997 to 2002.

values of NSE at the monthly and daily time scales, compared to unsatisfactory values for all the other approaches. The largest storm on record for the Little Washita, which occurred on 19-20 October 1983, significantly influenced the computation of NSE values for the wet climatic period. For that storm, SSQauto11 overestimated the daily peak by only 2%, compared to 58% for the manual, 63% for the SSQauto6, and 86% for the SSQRauto6 approaches.

Figure 6 compares monthly streamflow on Little Washita River subwatershed 526 for the calibration period from 1993 to 1999. The figure illustrates that the manual calibration generally produced a closer match with the observed streamflow in comparison to the SSQauto6 or SSQauto11 approaches. These three calibration approaches were considerably better than SSQRauto6, which substantially underestimated streamflow for the period from January 1993 to June 1993 and overestimated flows for the June 1996 to December 1996 period. A comparison of daily streamflow for subwatershed 526 from 1 January 1998 to 30 June 1998 shows that the manual calibration tended to overestimate some of the peak storms but for the most part simulated various portions of the hydrograph with reasonable accuracy (fig. 7). The SSQauto6 and SSQauto11 methods, on the other hand, underestimated most of the peak events and overestimated the recession side of the hydrographs on the subwatershed. The SSQRauto6 method produced similar results in estimating peak events, as did the other autocalibration approaches, but did a better job in simulating the recession side of the hydrograph. Figure 8 illustrates that for the most part, the manual approach outperformed the autocalibration ap−

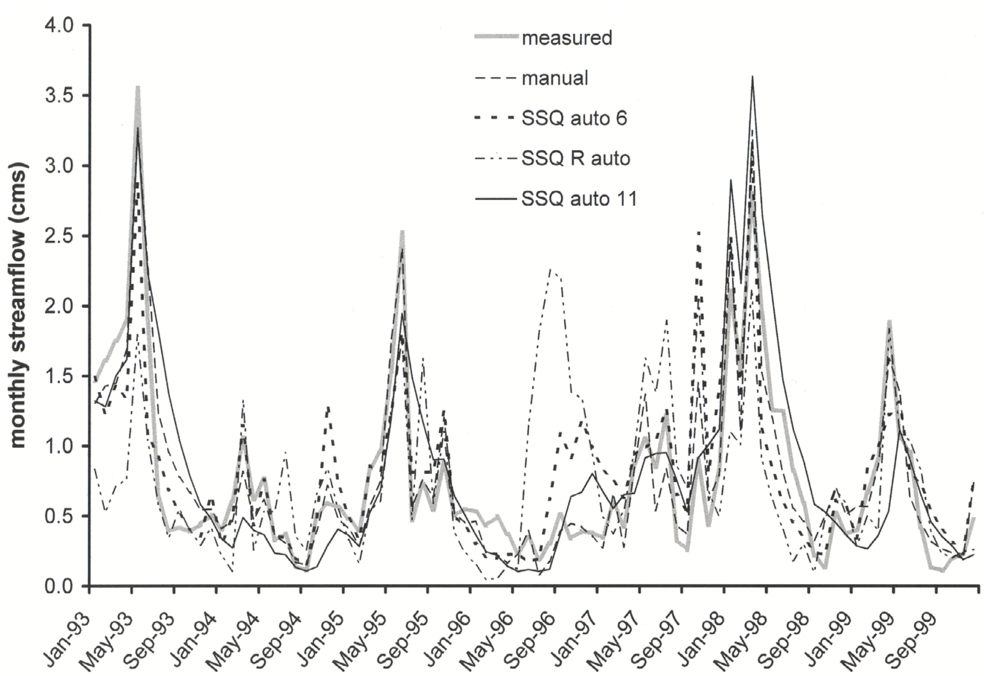

Figure 6. Comparison of monthly streamflow on Little Washita River subwatershed 526 for the calibration period from 1993 to 1999.

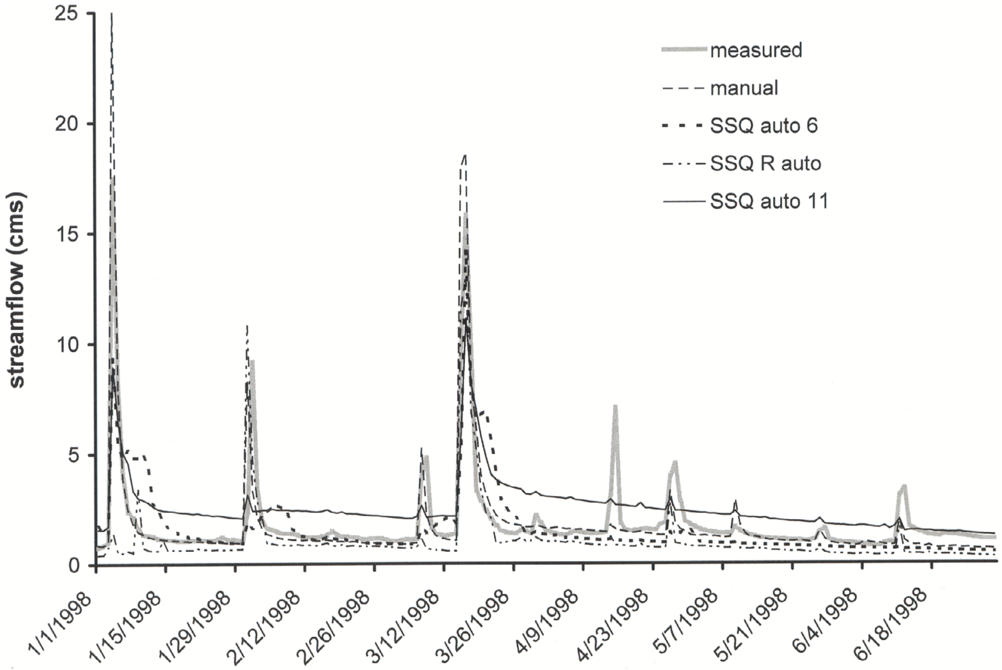

Figure 7. Comparison of daily flow for Little Washita River subwatershed 526 from 1 January 1998 to 30 June 1998.

proaches in simulating monthly streamflow on subwatershed 522 for the average climatic condition from 1975 to 1979. For this validation period, the manual method overestimated streamflow by only 8%, compared to 42%, 61%, and 27%, respectively, for the SSQauto6, SSQRauto6, and SSQauto11 methods. A comparison of duration of daily flow on subwatershed 550 for the calibration period shows that the SSQRauto6 method performed slightly better than the other methods in simulating the range in magnitude of the flows (fig. 9). Test results obtained on the Little Washita illustrate that the autocalibration approaches gave much more consistent results in estimating the lower flow ranges than were estimated with the autocalibration approaches on the Little River.

Model Performance Evaluation on an Annual Basis

PBIAS and NSE error statistics were used to evaluate model performance with the four calibration approaches on an annual basis (table 5). These two statistics were computed at the daily time step for each year of record. Listed in the table is the total number of years that model simulations were

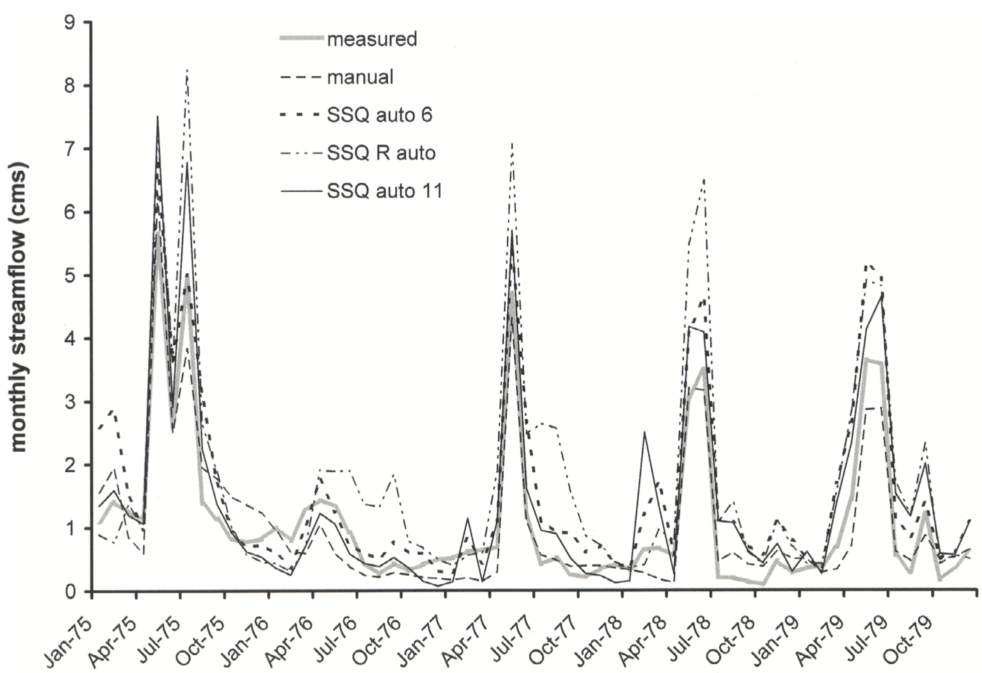

Figure 8. Comparison of monthly streamflow on Little Washita River subwatershed 522 for the "average climatic condition" validation period from 1975 to 1979.

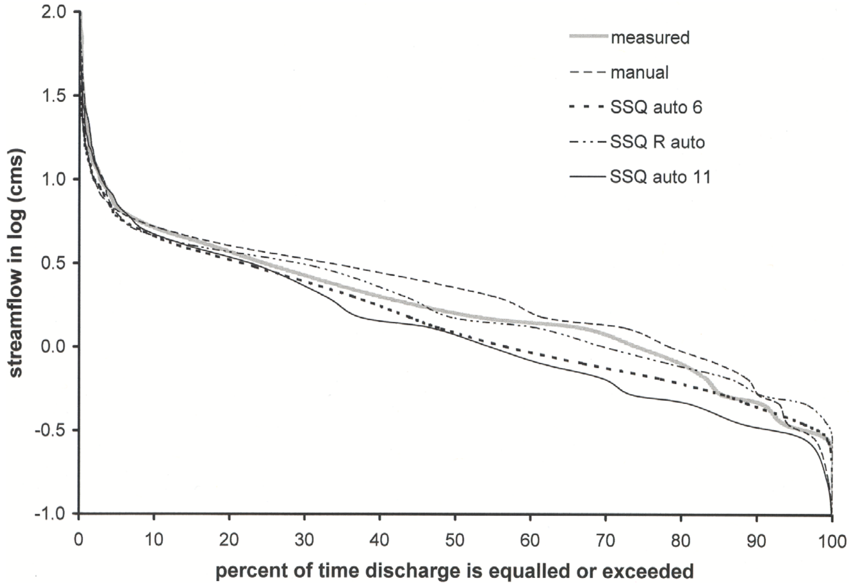

Figure 9. Duration of daily flow for the Little Washita subwatershed 550 calibration period from 1993 to 1999.

performed for a given watershed. Thus, PBIAS and NSE values were computed for 31 years of record on the Little River for the F and B subwatersheds. Thirteen years of record were simulated on Little Washita subwatershed 526, and available data for subwatersheds 522 and 550 were combined to give a 29-year period of record. Also listed in table 5 is the respective number of years that were considered good, satisfactory, and unsatisfactory, based on computed values of the two error statistics.

Based on the computation of NSE, test results show that for both subwatersheds of the Little River, SSQauto6 and SSQauto11 clearly outperformed the manual and SSQRauto6 calibrations. For the Little River subwatershed F dataset, 1 year simulated by the SSQauto6 approach was considered good, 24 were considered satisfactory, and 6 were considered unsatisfactory. This compares to 5 years that were considered satisfactory and 26 that were considered unsatisfactory for the manual calibration. The SSQauto11 approach gave slightly better results than the SSQauto6 approach for both subwatersheds.

Greater differences among calibration approaches were reflected in results obtained in comparing PBIAS. With 17 years considered good, 7 satisfactory, and 7 unsatisfactory, the manual calibration outperformed the automatic calibration approaches on Little River subwatershed F. For subwatershed B, the best results were obtained with the SSQauto11 approach, followed by the manual, SSQauto6, and SSQRauto6 approaches, respectively.

Test results indicate that the manual calibration for the most part outperformed the autocalibration approaches on the two subwatersheds of the Little Washita (table 5). Based on NSE computations, 2 years were considered good with the manual calibration, 8 were considered satisfactory, and 3 were considered unsatisfactory compared to 1, 7, and 5, respectively, with the SSQauto6 approach on subwatershed 526. For the combined 522/550 dataset, 1 year was considered good for the manual calibration, 14 were considered satisfactory, and 14 were considered unsatisfactory based on the NSE statistic. This compares to 1, 10, and 18, respectively, for SSQauto6; 1, 15, and 13 for SSQauto11; and 0, 1, and 28 for SSQRauto6. Test results also show that manual calibration outperformed the autocalibration methods on the Little Washita based on the PBIAS statistic. On subwatershed 526, 9 years were considered good, 4 satisfactory, and 0 unsatisfactory with the manual calibration compared to 6, 4, and 3 for SSQauto6 and 8, 2, and 3 for SSQauto11, respectively. PBIAS results show that on the combined 522/550 dataset, 18 years were considered good with the manual calibration compared to only 5, 5, and 7 years that were considered good for the SSQauto6, SSQRauto6, and SSQauto11 calibrations, respectively.

Table 5. Comparison of calibrated model performance on an annual basis using the percent bias (PBIAS) and Nash-Sutcliffe coefficient of efficiency (NSE) error statistics.

		PBIAS[a]			NSE[a]		
Calibration Method	No. of Years	More than ±40%	20% to 40%	Less than ±20%	Less than 0.36	0.36 to 0.75	More than 0.75
Little River F							
Manual	31	7	7	17	26	5	0
SSQRauto6		12	10	9	26	5	0
SSQauto6		9	12	10	6	24	1
SSQauto11		2	16	13	3	25	3
Little River B							
Manual	31	5	8	18	31	0	0
SSQRauto6		6	16	9	31	0	0
SSQauto6		10	4	17	6	18	7
SSQauto11		3	3	25	1	26	4
Little Washita 526							
Manual	13	0	4	9	3	8	2
SSQRauto6		4	4	5	10	3	0
SSQauto6		3	4	6	5	7	1
SSQauto11		3	2	8	10	3	0
Little Washita 522/550							
Manual	29	7	4	18	14	14	1
SSQRauto6		20	4	5	28	1	0
SSQauto6		21	3	5	18	10	1
SSQauto11		10	12	7	13	15	1

[a] Statistics computed at the daily time step.

DISCUSSION

The modeling experiences conducted in this study represent two approaches that can be undertaken in model calibration. These experiences accentuate the strengths and weaknesses of manual and autocalibration approaches in calibrating a watershed-scale simulation model such as SWAT. A distinct disadvantage that became apparent in the procedure to manually calibrate SWAT was the enormous amount of time required to adjust model parameters so that measured and simulated hydrologic responses agreed well. Approximately four to six weeks of labor were required for manually calibrating one of the watersheds in this study, compared to only one day using the autocalibration tool. Moreover, the manual approach was often fraught with a considerable degree of frustration and uncertainty as decisions were made during calibration. Labor-intensive efforts related to manual calibration were the result of the need to become familiar with the way that the model simulated responses on the two watersheds, to identify the sensitivity of the model to changes in parameter values, and to understand the nature of parameter interactions on hydrologic response. Although a significant advantage of the autocalibration tool in SWAT was that very little labor was required to conduct the parameter search, this approach suffered from the inability to maintain control on mass balance and in some cases to adequately represent the range in magnitude of daily flows. Model calibration on the two watersheds showed that when using the manual approach, it was possible to verify that these two facets of watershed response were preserved each time a run was made during calibration.

Of the parameters calibrated in this investigation, we found that values obtained in the runoff curve number (CN2) parameter search produced the most striking results. In all cases but two, CN2 values calibrated on the two watersheds exhibited departures from the default values that ranged from −13% to a staggering −50%. This tendency was more pronounced on the Little River watershed, where departures ranged from −27% to −50% from the default values. Percent departures computed in the autocalibration parameter search were similar to a departure of −27% determined in the manual approach, and indicate that greater contributions of precipitation to subsurface than surface runoff (as reflected by low curve numbers) result in a closer match between measured and simulated streamflow. Such low values of the curve number that were used in the simulation on the Little River seem unrealistic, and more than likely reflect a model deficiency in using this approach to adequately account for physical processes that occur in the precipitation runoff partitioning mechanism on the watershed. Limiting the potential range of the curve number to ±10% from the model default values appears to be a much more plausible approach to model calibration. Limiting the range in values of the surface runoff lag time (SURLAG) should also be considered when calibrating this parameter in SWAT. Cursory testing with a 0.0 to 10.0 range for SURLAG led to improved model performance on both the Little River and Little Washita watersheds. Although selection of a lower limit for SURLAG is somewhat arbitrary, we believe that parameter values less than 0.5 do not properly account for the release of surface runoff from a subbasin to the main channel. Whether manual or automated calibration is used, limiting the initial range in values for parameters such as the curve number and the surface runoff lag time in SWAT is important in minimizing model distortions of hydrologic processes occurring on the watershed.

Based on PBIAS, simulation results from this study indicate that the SSQauto6 method outperformed the SSQRauto6 method on the Little River watershed, but that both methods performed about equally well on the Little Washita. In nearly all cases, SSQauto11 gave much better results than the other two autocalibration approaches on both watersheds, based on PBIAS. Results of this investigation show that in almost all cases, the SSQauto6 approach outperformed the SSQRauto6 approach, based on NSE. This was not surprising, since the SSQ approach is biased towards the use of the NSE statistic, as mentioned earlier. Because the SSQ minimized differences between storm peaks rather than the differences between ranked flows, which were minimized using the SSQR approach, the SSQ optimization scheme yielded a streamflow response that more closely simulated storm hydrographs. The SSQR scheme did a somewhat better job estimating the range in magnitude of the daily flows. Because the parameter search in both the SSQ and SSQR optimization schemes was based only on minimizing differences between measured and simulated variables, there was no control in maintaining mass balance for respective periods used in model calibration. Results of the calibration (1997 to 2002) of the Little River using the SSQRauto6 well illustrate this problem: streamflow was overestimated by 24% for subwatershed F and underestimated by 19% for subwatershed B. Improved model performance that was obtained with the SSQauto11 approach in comparison to SSQauto6 or SSQRauto6 was mainly evidenced by daily NSE values for the Little River watershed and daily NSE and PBIAS values for the three validation periods representing varying climatic conditions on Little Washita subwatershed 522. Better model performance using the SSQauto11 reflects the fact that with the additional five parameters that were calibrated, a wider range in options was available for the model to match measured versus simulated streamflow responses. Results obtained from the SSQauto11 calibration suggest that for the most part SWAT exhibits robustness in simulating streamflow responses from agricultural watersheds with very different features. Results obtained in using the SSQauto11 calibration are supportive of previous studies that demonstrated the versatility of the model and its predecessor SWRRB in simulating hydrologic responses from agricultural watersheds under a range in climatic, soils, and land use conditions (Van Liew et al., 2003; Borah and Bera, 2003; Arnold and Williams, 1987).

This study highlights an important difference that must be realized in comparing the manual versus autocalibration approaches that were used on the two watersheds. The autocalibration approach was strictly a quantitative comparison that involved minimizing the difference between measured and simulated values (eqs. 1 or 2). The manual approach involved both quantitative and qualitative comparisons, since it involved using the total mass controller (eq. 3) in conjunction with graphical comparisons of monthly and daily hydrographs and duration of daily flow curves to calibrate the model against measured data. Use of the manual calibration accentuates the tradeoffs that exist in achieving total mass balance, reasonable hydrograph responses, and adequate representation of the range in flows. It may be said, therefore, that the governing approach taken to calibrate a given watershed depends on the particular needs that must be addressed in modeling. For example, short-duration storm runoff analysis, low flow assessments, water availability evaluations, and the impact of climate variations on streamflow all represent different types of problems that will dictate how a given calibration should be performed. Findings from this study suggest that the strengths of both the manual and autocalibration approaches can be utilized to facilitate the calibration process. With the proper selection of the upper and lower ranges in values for model parameters, autocalibration can be used to provide an initial parameter set with minimal labor on the part of the practitioner. Depending on the particular modeling needs, a manual approach can then be taken to refine the calibration, so that an appropriate balance is achieved regarding the amount, timing, and distribution of the output variable.

Conclusions and Recommendations

In this study, an investigation was conducted to evaluate manual (or expert) calibration and autocalibration approaches in SWAT. Manual calibrations performed on the Little River and Little Washita watersheds reflect much greater variability than were realized in the automated approaches. Differences in the comparison of manual results on the two watersheds are mainly attributed to two factors. One factor is that a particularly suitable set of parameters could not be found to adequately account for the wide variations in streamflow response during the calibration period on the Little River watershed. The other factor is that the presence of the flood-retarding structures that attenuate flows on the Little Washita likely led to improved performance of the model, since a somewhat narrower range in flows was simulated by the model from one day to the next during storm events on the watershed.

Test results show that the manual calibration generally outperformed the autocalibration tool in the estimate of percent bias and simulation of the range in magnitude of daily flows. Based on the coefficient of efficiency, the six-parameter and eleven-parameter autocalibration approaches using the sum of squares of residuals objective function gave significantly better results than did the manual calibration on the Little River watershed. Differences in model performance between the manual and automated approaches implemented in this study reflect the outcome of possible options that are selected for the objective function or qualitative criteria used in calibration. How the approach to calibration is selected is therefore dependent on the nature of the water resources problem to be addressed on the watershed.

Results of this study suggest that the autocalibration option in SWAT provides a powerful, labor-saving tool that can be used to substantially reduce the frustration and uncertainty that often characterizes manual calibrations. If used in combination with a manual approach, the autocalibration tool shows promising results in providing initial estimates for model parameters. To maintain mass balance and adequately represent the range in magnitude of output variables, manual adjustments may be necessary following autocalibration. Caution must also be exercised in utilizing the autocalibration tool so that the selection of initial lower and upper ranges in the parameters results in calibrated values that are representative of watershed conditions.

References

Allen, P. B., and J. W. Naney. 1991. Hydrology of the Little Washita River watershed, Oklahoma. USDA-ARS Report ARS-90. Washington, D.C.: USDA Agricultural Research Service.

Arnold, J. G., and P. M. Allen. 1996. Estimating hydrologic budgets for three Illinois watersheds. *J. Hydrology* 176: 57-77.

Arnold, J. G., and J. R. Williams. 1987. Validation of SWRRB-simulator for water resources in rural basins. *J. Water Resources Planning* 113(2): 243-256.

Arnold, J. G., P. M. Allen, and G. Bernhardt. 1993. A comprehensive surface-groundwater flow model. *J. Hydrology* 142: 47-69.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large-area hydrologic modeling and assessment: Part I. Model development. *J. American Water Resources Assoc.* 34(1): 73-89.

Borah, D. K., and M. Bera. 2003. Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Trans. ASAE* 46(6): 1553-1566.

Bosch, D. D., J. M. Sheridan, H. L. Batten, and J. G. Arnold. 2004. Evaluation of the SWAT model on a coastal plain agricultural watershed. *Trans. ASAE* 47(5): 1493-1506.

Boyle, D. P., H. V. Gupta, and S. Sorooshian. 2000. Toward improved streamflow forecasts: Value of semidistributed modeling. *Water Resources Research* 37(11): 2749-2759.

Bultot, F., and G. L. Dupriez. 1976. Conceptual hydrological model for an average-sized catchment area: I. Concept and relationships. *J. Hydrology* 29(3-4): 251-292.

Carrubba, L. 2000. Hydrologic modeling at the watershed scale using NPSM. *J. American Water Resources Assoc.* 36(6): 1237-1246.

Duan, Q. D. 2003. Global optimization for watershed model calibration. In *Calibration of Watershed Models*, vol. 6: 89-104. Water Sci. Appl. Series. Q. Duan et al., eds. Washington, D.C.: AGU.

Duan, Q. D., V. K. Gupta, and S. Sorooshian. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28(4): 1015-1031.

Green, W. H., and G. A. Ampt. 1911. Studies on soil physics, 1. The flow of air and water through soils. *J. Agric. Sci.* 4: 11-24.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.* 4(2): 135-143.

Hargreaves, G. H. 1975. Moisture availability and crop production. *Trans. ASAE* 18(5): 980-984.

Jacomino, V. M. F., and D. E. Fields. 1997. A critical approach to the calibration of a watershed model. *J. American Water Resources Assoc.* 33(1): 143-154.

Johanson, R. C., J. C. Imhoff, J. L. Little, and A. S. Donigian. 1984. User's manual for the Hydrological Simulation Program (HSPF). Athens, Ga.: U.S. Environmental Protection Agency.

Kellogg, R. L., M. S. Maizel, and D. W. Goss. 1994. The potential for leaching of agrichemicals used in crop production: A national perspective. *J. Soil and Water Cons.* 49(3): 294-298.

Knisel, W. G. 1980. CREAMS, a field-scale model for chemicals, runoff, and erosion from agricultural management systems. Conservation Research Report No. 26. Washington, D.C.: USDA.

Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Sindon. 1983. Precipitation-runoff modeling system: User's manual. USGS Water Resource Investment Report 83-4238. Washington, D.C.: U.S. Geological Survey.

Leonard, R. A., W. G. Knisel, and D. A. Still. 1987. GLEAMS: Groundwater loading effects on agricultural management systems. *Trans. ASAE* 30(5): 1403-1428.
Lowrance, R. R., J. K. Sharpe, and J. M. Sheridan. 1986. Long-term sediment deposition in the riparian zone of a Coastal Plain watershed. *J. Soil Water Conserv.* 42(4): 266-271.
Motovilov, Y. G., L. Gottschalk, K. Engeland, and A. Rodhe. 1999. Validation of distributed hydrological model against spatial observations. *Agric. and Forest Meteorology* 98: 257-277.
Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrology* 10(3): 282-290.
Neitsch, S. L., A. G. Arnold, J. R. Kiniry, J. R. Srinivasan, and J. R. Williams. 2002. Soil and water assessment tool user's manual: Version 2000. GSWRL Report 02-02, BRC Report 02-06. College Station, Texas: Texas Water Resources Institute TR-192.
Nelder, J. A., and R. A. Mead. 1965. Simplex method for function minimization. *Computer J.* 7(4): 308-313.
Sheridan, J. M. 1997. Rainfall-streamflow relations for coastal plain watersheds. *Trans. ASAE* 13(3): 333-344.
Srinivasan, M. S., J. M. Hamlett, R. L. Day, J. I. Sams, and G. W. Petersen. 1998. Hydrologic modeling of two glaciated watersheds in northeast Pennsylvania. *J. American Water Resources Assoc.* 34(4): 963-978.
Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian. 2001. Bayesian recursive parameter estimation for hydrologic models. *Water Resources Res.* 37(10): 2521-2535.
Van Griensven, A. 2002. Developments towards integrated water quality modeling for river basins. Publication No. 40. Brussels, Belgium: Vrije Universiteit, Department of Hydrology and Hydraulic Engineering.
Van Griensven, A., and W. Bauwens. 2003. Multiobjective autocalibration for semidistributed water quality models. *Water Resources Res.* 39(12): 1348.
Van Liew, M. W., and J. Garbrecht. 2003. Hydrologic simulation of the Little Washita River experimental watershed using SWAT. *J. American Water Resources Assoc.* 39(2): 413-426.
Van Liew, M. W., J. G. Arnold, and J. D. Garbrecht. 2003. Hydrologic simulation on agricultural watersheds: Choosing between two models. *Trans ASAE* 46(6): 1539-1551.
Water Quality and Watershed Research Laboratory. 1983. Hydrology, erosion, and water quality studies in the Southern Great Plains research watershed, southwestern Oklahoma, 1961-1978. USDA-ARS ARM-S-29. Chickasha, Okla.: USDA-ARS Water Quality and Watershed Research Laboratory.
Williams, J. R., A. D. Nicks, and J. G. Arnold. 1985. Simulator for water resources in rural basins. *J. Hydraulic Eng.* 111(6): 970-986.
Williams, J. R., C. A. Jones, and P. T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity. *Trans. ASAE* 27(1): 129-144.